

- 17 -

Claims

1. A method of extracting information from a natural-language text, by identifying patterns, said method being characterized in that the words of the text are encoded by comparing them with the contents of a predefined lexicon containing a few tens of tool words, and in that nominal groups are then identified by searching subsets of the resulting succession of encoded words to look for groups of encoded words that comply with predefined syntactical rules.

2. A method according to claim 1, characterized in that the words of the text are encoded by evaluating the grammatical function of each word by comparing it with the contents of said lexicon of tool words, so as to identify the tool words in the text, and in that the functions of the usage words, which are not recognized as being tool words, are deduced by comparing their locations relative to the locations of the words recognized as being tool words.

3. A method according to claim 1 or claim 2, characterized in that the identified nominal groups are then evaluated so as to keep only those groups which are perceived as being the most important, by using predefined evaluation criteria.

4. A system for extracting information from a natural-language text, said system being characterized in that it comprises:

- an input unit for receiving said natural-language text;
- a lexicon file in which tool words are recorded;
- an analysis processor connected to said input unit, and to the lexicon file, and organized to act in a first stage to encode the words of the natural-language text by evaluating the grammatical function of each word by comparing it with the contents of said lexicon file of tool words, so as to identify the tool words in the text and so as to evaluate the functions

- 18 -

of the usage words which are not recognized as being tool words, by comparing their locations relative to the locations of the words recognized as being tool words, and, in a second stage, to search subsets of the resulting succession of encoded words to look for groups of encoded words that comply with predefined syntactical rules, so as to identify nominal groups; and

- an output unit connected to said analysis processor for receiving the groups of encoded words recognized as being syntactical patterns.

5. A system according to claim 4, characterized in that the analysis processor further comprises means for evaluating the importance of the kept groups of encoded words in order to keep only those groups which are perceived as being the most important.

6. A system according to claim 3 or claim 4, characterized in that the analysis processor further comprises means for recognizing the language of the text received at the input unit.

7. A system according to any one of claims 4 to 6, characterized in that the analysis processor further comprises means for regularizing the text received at the input unit so as to remove the amalgams of signs.